REAL-TIME FORECASTING OF COVID-19 PATIENT SEVERITY USING BIG DATA ANALYTICS

Don Vindhya Dilini Kalapuge

A dissertation submitted in partial fulfilment of the requirement for Master of Science degree in Big Data Analytics

Department of Computing
Informatics Institute of Technology, Sri Lanka
in collaboration with
Robert Gordon University, UK

2022

Abstract

COVID-19 has been declared a global health emergency by the World Health Organization March 11, 2020. From the date of the first confirmed case in China on November 4, 2021, there have been over two hundred million confirmed cases worldwide, with more than five million deaths. Many countries initially recorded many casualties due to a shortage of healthcare resources. The presence of adequate facilities and medical staff in hospitals is most beneficial in reducing the mortality rate. The number of patients hospitalized can be managed by considering the intensity of their symptoms as well as other prevailing medical conditions. The World Health Organization discovered that the severity of a patient's risk level is dependent on their symptoms, commodities, and some demographic characteristics which would be crucial in managing the hospital resources.

The application of information technology, with a focus on subjects such as data science and machine learning, can contribute to the fight against the epidemic. The proposed forecasting application with machine learning would allow hospital administration to detect patients' severity levels based on their symptoms, medical circumstances as well as demographic characteristics. The severity forecasting tool is a user-friendly web application built with a machine learning model where the data was obtained from Brazilian Health Ministry.

The data is obtained through streaming applications using big data tools and Spark platform was used to incorporate Spark machine learning libraries to produce machine learning models. By using Multi-Layer Perceptron Classifier, a 60% of accuracy rate was obtained. All available data is stored in a database, which can be queried by authenticated user.